



FRANCE STRATÉGIE

ÉVALUER ANTICIPER DÉBATTRE PROPOSER

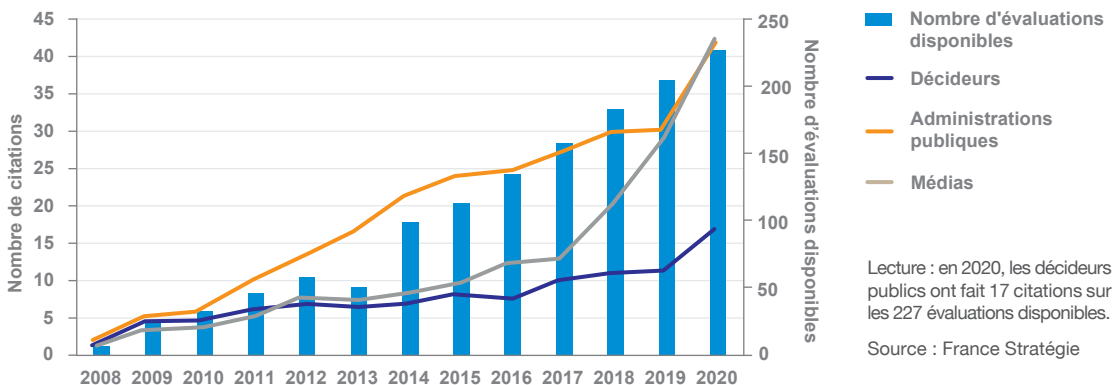
# Qui utilise les évaluations académiques des politiques publiques ?

En matière d'évaluations académiques sur l'impact des politiques publiques, la France rattrape son retard par rapport aux pays les plus avancés<sup>1</sup>. Ces évaluations sont-elles pour autant davantage utilisées dans le débat public ? Pour explorer cette question, on considère ici exclusivement les évaluations quantitatives et causales publiées dans les revues académiques à comité de lecture et portant de façon *ex post* sur le critère évaluatif d'impact. En partant d'un échantillon de 227 évaluations, l'analyse recense ainsi près de 8 300 citations sur la période 2008-2020<sup>2</sup>.

Premier constat, les citations sont faites à une écrasante majorité par des chercheurs (à 94 %). Les décideurs publics, les administrations et les médias y ont peu recours : 60 évaluations de notre échantillon n'ont même jamais été citées par ces acteurs sur la période. Second constat : de façon agrégée, les citations augmentent rapidement sur la période 2012-2020, passant de 230 citations en 2012 à près de 1 700 en 2020. La hausse est plus rapide que celle du nombre d'évaluations disponibles : ainsi chaque évaluation disponible est citée 3,9 fois en 2012, contre 7,4 fois en 2020. Pour les décideurs publics, les administrations publiques et les médias réunis, le nombre de citations est multiplié par 3,6 entre 2012 et 2020. Toutefois, ce nombre reste stable lorsque l'effet volume est corrigé (0,4 citation par évaluation disponible).

La thématique « économie » est beaucoup évaluée (30 fois) et ses évaluations sont parmi les plus citées, avec plus de 50 citations en moyenne par évaluation. Par contraste, des thématiques comme « sécurité et police » et « culture » donnent lieu à peu d'évaluations, qui figurent parmi les moins citées (entre 0 et 6 citations par évaluation). Par ailleurs, les rares évaluations sur les thématiques « famille » et « environnement » (6 évaluations chacune) sont davantage citées que les très nombreuses évaluations sur la thématique « travail » : on relève respectivement 53, 42 et 35 citations en moyenne par évaluation. Toutes thématiques confondues, les évaluations sont d'autant plus citées par ces acteurs qu'elles sont en français, publiées par des chercheurs issus d'une administration publique ou qu'elles formulent des résultats riches – pas nécessairement conclusifs – et des recommandations suffisamment larges. Confirmés par des entretiens auprès de chercheurs et de commissions de l'Assemblée nationale, ces faits stylisés pourraient inspirer des bonnes pratiques à tout évaluateur soucieux d'éclairer le débat public.

## Nombre de citations faites annuellement 2008-2020



Lecture : en 2020, les décideurs publics ont fait 17 citations sur les 227 évaluations disponibles.

Source : France Stratégie

1. Bono P.-H. et al. (2018), « Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique », Document de travail no 2018-06, France Stratégie, décembre.  
2. Pour le détail des méthodes et des résultats, voir Baïz A., Guyot M., Lewandowski M. et Suty A. (2022), *Quelles évaluations des politiques publiques pour quelles utilisations ?*, Rapport, France Stratégie, juin. Pour une étude fondée sur une définition plus large de l'évaluation, voir la seconde note issue de ce rapport : Baïz A., Guyot M., Lewandowski M. et Suty A. (2022), « Quelles évaluations sont mobilisées avant et après le vote d'une loi ? », *La Note d'analyse*, n° 110, France Stratégie, juin.

Adam Baïz

avec la collaboration de  
Mathilde Guyot,  
Marianne Lewandowski,  
Achille Suty

département Économie

La *Note d'analyse* est publiée sous la responsabilité éditoriale du commissaire général de France Stratégie. Les opinions exprimées engagent leurs auteurs et n'ont pas vocation à refléter la position du gouvernement.

## INTRODUCTION GÉNÉRALE

Plusieurs rapports institutionnels<sup>3</sup> font état des obstacles à l'utilisation des travaux évaluatifs par les décideurs publics et dans le débat public. Ces obstacles concernent les potentiels utilisateurs des évaluations (un défaut d'acculturation par exemple), les évaluations elles-mêmes (un défaut d'intelligibilité par exemple), et de façon évidente l'indisponibilité des évaluations aux moments opportuns. Les recommandations formulées dans ces rapports tentent de répondre à ces difficultés.

La littérature académique s'est également intéressée à l'utilité théorique et aux usages pratiques de l'évaluation dans la conduite de l'action publique<sup>4</sup>. Elle fait état de la multiplicité des usages. Ainsi, une évaluation peut être utilisée par une administration afin d'améliorer la conduite d'une politique publique (usage instrumental) ; elle peut être utilisée par un chercheur ou un citoyen pour acquérir des connaissances et mieux comprendre les tenants et les aboutissants de l'intervention publique (usage conceptuel ou cognitif) ; elle peut aussi être utilisée par un décideur public cherchant à justifier la pertinence de ses décisions (usage persuasif, voire symbolique). En particulier, l'évaluation peut servir à travers les connaissances qu'elle produit. Notamment, elle peut renseigner sur les coûts d'une intervention. L'évaluation peut aussi être utile à travers le processus même de sa réalisation. Elle peut par exemple associer plusieurs parties prenantes et ainsi faciliter la mise en œuvre de ses conclusions. En outre, les évaluations peuvent être également mal utilisées, par méconnaissance des sujets, par opportunisme politique, voire par mauvaise foi. Elles peuvent également être ignorées.

Dans une approche empirique, quelques travaux appréhendent l'utilisation des évaluations à travers des études de cas. Rares sont ceux qui proposent une analyse plus systématique et sur de larges échantillons. La méta-évaluation réalisée par KPMG et Quadrant Conseil en 2017 autour de l'utilisation de 65 évaluations relatives à la Modernisation de l'action publique est un rare exemple disponible<sup>5</sup>. Tous ces travaux ont tendance à souligner la faible utilisation réelle des évaluations dans le processus de décision politique. La présente analyse propose de vérifier ce constat à partir de l'ensemble des évaluations d'impact publiées dans les revues académiques entre 2008 et 2020.

Grâce à une analyse textuelle inédite, sont construits plusieurs indicateurs quantitatifs. En découle une analyse statistique au croisement de l'utilisation des évaluations d'impact et des caractéristiques de ces dernières. Outre l'analyse textuelle, une centaine d'échanges et d'entretiens ont également été réalisés auprès d'évaluateurs et de commissions parlementaires afin d'affûter les hypothèses et d'étayer les analyses statistiques.

## UN ÉCHANTILLON DE 227 ÉVALUATIONS ACADÉMIQUES D'IMPACT PUBLIÉES ENTRE 2008 ET 2020

À partir de la base EconLit, les évaluations d'impact sont identifiées à travers les mêmes mots clés (*random experiment, matching*, etc.) que dans l'étude bibliométrique de 2018. Les évaluations doivent vérifier quatre caractéristiques : (a) porter sur des politiques publiques françaises ; (b) interroger le critère d'impact de façon *ex post* ; (c) reposer sur un raisonnement de type contrefactuel (expérimentation aléatoire, *matching*, régression sur discontinuité, doubles différences, etc.) ; (d) avoir été publiées dans une revue scientifique à comité de lecture. Afin de constituer un échantillon exhaustif et exact, un ensemble de vérifications humaines a été réalisé : lecture des articles, échanges avec leurs auteurs et mobilisation de moteurs de recherche alternatifs. Ces différentes corrections ont permis d'aboutir à un total de 227 évaluations.

### *Les utilisateurs potentiels des évaluations : les décideurs, les administrations, les médias et les chercheurs*

L'utilisation des 227 évaluations a été mesurée à travers le nombre de citations, distinguées selon la nature des acteurs qui sont susceptibles de les faire :

- **Les citations des décideurs publics** (gouvernement et Parlement), dans le cadre de rapports d'information, de communiqués de presse ou encore de rapports commandités par le gouvernement. Par exemple, l'évaluation de Cochard *et al.* (2008)<sup>6</sup> est citée par le *Rapport n° 159 fait au nom de la commission des finances sur le projet de loi de finances rectificative (2014-2015)* de M. Albéric de Montgolfier, déposé le 10 décembre 2014 à la présidence du Sénat.

3. Voir notamment : Cese (2015), « *Voies possibles d'une promotion de la culture de l'évaluation* » ; Assemblée nationale (2018), *Rapport d'information du CEC sur l'évaluation des dispositifs d'évaluation des politiques publiques*, Comité d'évaluation et de contrôle des politiques publiques ; Desplatz R. et Lacouette-Fougère C. (2019), « *L'évaluation des politiques publiques en France* », Document de travail no 2019-13, France Stratégie, décembre ; ou encore : Conseil d'État (2020), *Conduire et partager l'évaluation des politiques publiques*, Étude annuelle 2020, Paris, La documentation française.

4. En particulier, deux méta-analyses méritent d'être mentionnées : Delarue J. *et al.* (2009), « *Les évaluations sont-elles utiles ?* », Série Notes méthodologiques, Agence française de développement, janvier ; et, plus récemment, T. Delahais *et al.* (2022), *Évaluation : Fondements, controverses, perspectives*, anthologie, Sciences Po.

5. SGMAP (2017), *Évaluation de la démarche globale d'évaluation des politiques publiques menée dans le cadre de la modernisation de l'action publique*, rapport réalisé par KPMG et Quadrant Conseil, février.

6. Cochard M., Junod-Mesqui B., Arnaud F. et Vermare S. (2008), « *Les effets incitatifs de la prime pour l'emploi : une évaluation difficile* », *Économie et Statistique*, n° 412, octobre, p. 57-80.



- **Les citations des administrations ou des institutions publiques.** Les citations sont identifiées dans des productions d'organes de l'exécutif (Insee, services statistiques ministériels<sup>7</sup>, DITP, etc.), d'établissements publics (Pôle emploi, Ademe, AFD, etc.), ou encore d'organisations internationales (OCDE, ONU, OMC, OMS, etc.). Par exemple, l'évaluation de Behaghel *et al.* (2014)<sup>8</sup> est citée dans le rapport de la Cour des comptes *Pôle emploi à l'épreuve du chômage de masse* (juillet 2015).
- **Les citations des chercheurs entre eux.** Par exemple, l'évaluation de Gurgand *et al.* (2017)<sup>9</sup> est citée 63 fois par des travaux de recherche, dont l'article de Beuchert *et al.* (2018)<sup>10</sup>.
- **Les citations dans les médias** (presse, émissions de radio retranscrites, etc.). Par exemple, l'évaluation de Lequien *et al.* (2012)<sup>11</sup> est citée par l'article « Le congé parental nuit plus au salaire qu'à la carrière des femmes » de *Capital* (juillet 2012).

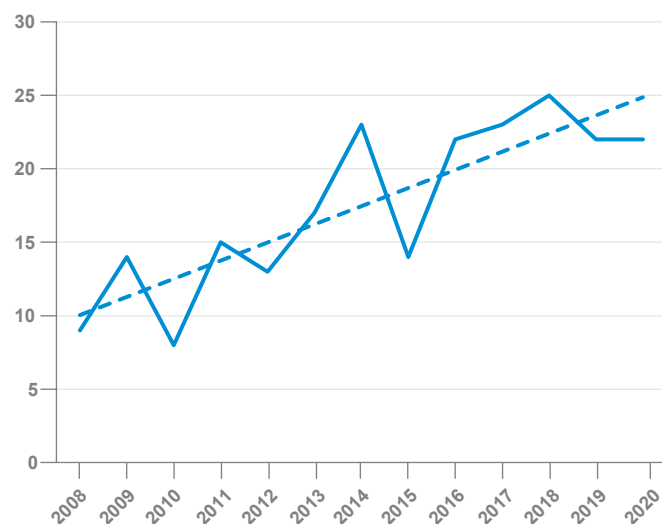
### Présentation de l'échantillon au regard des facteurs susceptibles de favoriser l'utilisation des évaluations d'impact

En remarquant que les évaluations sont plus ou moins citées, en général et selon les acteurs susceptibles de les

citer, il a été question de comparer les caractéristiques des évaluations les plus citées à celles des évaluations les moins citées. Bien que cette comparaison relève davantage d'un calcul de corrélations que d'une démonstration causale, elle permet d'esquisser un ensemble de bonnes pratiques dont les évaluateurs pourraient s'inspirer pour que leurs travaux soient perçus comme plus utiles et soient *in fine* davantage utilisés dans le débat public. Au regard des entretiens réalisés et de la littérature académique, huit caractéristiques susceptibles de favoriser l'utilisation des évaluations d'impact ont été choisies :

- **L'année de publication de l'évaluation :** dans l'échantillon, le nombre d'évaluations publiées annuellement augmente tendanciellement sur la période 2008-2020, en passant d'une moyenne de 9 évaluations en 2008 à près de 25 en 2020.
- **La langue de la publication :** dans l'échantillon étudié, la plupart des évaluations sont rédigées en anglais. Sur les 227 évaluations, 147 sont rédigées en anglais contre 80 en français. La part des évaluations rédigées en français chute drastiquement dès 2014. Ces tendances tiennent vraisemblablement au plus grand nombre de revues académiques anglophones, et à leur prestige dans le monde de la recherche.

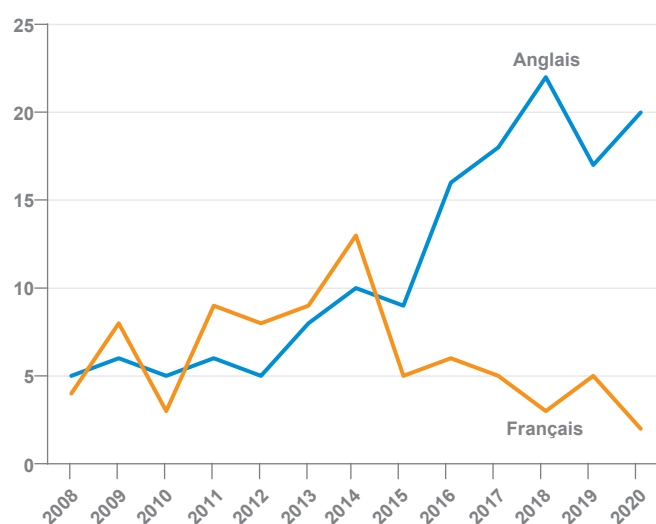
### Graphique 1 – Évolution du nombre d'évaluations d'impact



Lecture : en 2014, 23 évaluations d'impact ont été publiées.

Source : France Stratégie

### Graphique 2 – Langue des évaluations publiées



Lecture : en 2020, deux évaluations ont été publiées en français et vingt en anglais.

Source : France Stratégie

7. Si un ministère (ou plus généralement une administration) cite un article académique dans le cadre d'un rapport expressément commandité par le gouvernement, la citation sera classée du côté des « décideurs » ; ce rapport aura en effet plus de chances d'apparaître comme l'émanation directe d'une décision politique.

8. Behaghel L., Crépon B. et Gurgand M. (2014), « Private and public provision of counseling to job seekers: Evidence from a large controlled experiment », *American Economic Journal: Applied Economics*, vol. 6 (4), octobre, p. 142-174.

9. Behaghel L., de Chaisemartin C. et Gurgand M. (2017), « Ready for boarding? The effects of a boarding school for disadvantaged students », *American Economic Journal: Applied Economics*, vol. 9(1), janvier, p. 140-164.

10. Beuchert L., Humlum M. K., Nielsen H. S. et Smith N. (2018), « The short-term effects of school consolidation on student achievement: Evidence of disruption? », *Economics of Education Review*, no 65, p. 31-47.

11. Lequien L. (2012), « The impact of parental leave duration on later wages », *Annales d'Économie et de Statistique*, n° 107-108, juillet-décembre, p. 267-285.

- **La thématique de la politique évaluée** : dans l'échantillon, certaines thématiques<sup>12</sup> font particulièrement l'objet de travaux évaluatifs. En effet, 70 % de l'échantillon considéré concernent 4 thématiques, soit 159 évaluations sur un total de 227. Il s'agit des thématiques relatives au travail (61 évaluations), à l'éducation (38 évaluations), aux questions sociales et de santé (31 évaluations) et à l'économie (29 évaluations). Les autres thématiques sont relativement moins évaluées, et notamment les politiques liées à la recherche, à la sécurité et à la police, à la culture ou encore à l'environnement, avec moins de 6 évaluations pour chacune de ces thématiques.
- **La nature du producteur de l'évaluation** : dans l'échantillon, sont distingués cinq types de producteurs d'évaluation. Parmi les 227 évaluations, il ressort que les évaluations sont produites : (a) à 26 % par une unique université ou un unique centre de recherches<sup>13</sup> ; à 41 % par l'association de plusieurs universités ou centres de recherches<sup>14</sup> ; à 11 % par une ou plusieurs administrations publiques<sup>15</sup> ; à 21 % par l'association d'administrations publiques avec des universités ou des centres de recherches<sup>16</sup> ; et à 1 % par d'autres acteurs<sup>17</sup>.
- **La mobilisation ou non d'une analyse qualitative** : dans l'échantillon, l'ensemble des 227 évaluations

cherchent à apprécier l'impact d'une politique publique à partir de méthodes quantitatives. Parmi elles, 22 incluent également une analyse qualitative (entretien, sondage, observation du terrain, etc.), soit environ 10 % de l'échantillon<sup>18</sup>. Une grande majorité des évaluations d'impact recensées (plus de 90 %) reposent donc exclusivement sur des méthodes quantitatives.

- **Le recours ou non à une méthode d'expérimentation aléatoire** : dans l'échantillon, 7 % reposent sur une telle méthode, consistant à tirer au sort un groupe de contrôle et un groupe de traitement<sup>19</sup>. Les autres évaluations reposent sur d'autres méthodes contrefactuelles quantitatives (régression sur discontinuité, modélisation, *matching*, différence de différences, etc.).
- **La nature conclusive des résultats de l'évaluation** : en relevant la significativité des résultats statistiques, sont distinguées dans l'échantillon les évaluations conclusives (57 % de l'échantillon<sup>20</sup>), conditionnellement conclusives (12 % de l'échantillon<sup>21</sup>) et non conclusives (30 % de l'échantillon<sup>22</sup>).
- **La présence ou non de recommandations** : dans l'échantillon, 59 % des évaluations ne formulent aucune recommandation ; 8 % formulent des recommandations précises<sup>23</sup> ; et 33 % formulent des orientations générales<sup>24</sup>.

12. À partir de la classification JEL (pour *Journal of Economic Literature*), sont distinguées vingt thématiques, correspondant chacune à un champ de la politique publique : la justice ; l'économie ; le travail ; la société ; les enjeux de police et de sécurité ; les questions sociales et de santé ; la culture ; les transports ; l'environnement ; les pouvoirs publics ; l'éducation ; la fonction publique ; le logement et l'urbanisme ; la famille ; l'agriculture ; les sports ; les entreprises ; l'énergie ; le tourisme ; et la recherche.
13. Voir par exemple Paty *et al.* (2020), où les auteurs sont tous issus de l'université Lyon 2. Ly T. et Paty S. (2020), « [Local taxation and tax base mobility: Evidence from France](#) », *Regional Science and Urban Economics*, vol. 82, mai.
14. Voir par exemple Cahu *et al.* (2010), où les auteurs sont issus de Sciences Po, du centre allemand IZA et de l'université italienne Bocconi. Cahu P., Demmou L. et Massé E. (2010), « [L'impact macroéconomique de la réforme 2008 du crédit d'impôt recherche](#) », *Revue économique*, vol. 61(2), mars, p. 313-339.
15. Voir par exemple Aeberhardt *et al.* (2009), où les auteurs sont issus de l'Insee au moment de la publication. Aeberhardt R. et Sraer D. (2009), « [Allègements de cotisations patronales et dynamique salariale](#) », *Économie et Statistique*, n° 429-430, janvier, p. 177-189.
16. Voir par exemple Carloni *et al.* (2019), où les auteurs sont issus de l'université de Californie et du Congressional Budget Office américain. Benzarti Y. et Carloni D. (2019), « [Who really benefits from consumption tax cuts? Evidence from a jarge VAT reform in France](#) », *American Economic Journal: Economic Policy*, vol. 11(1), février, p. 38-63.
17. Voir par exemple Goupille-Lebret *et al.* (2016), où les auteurs sont issus de PSE et de la société d'assurance Axa. Goupille-Lebret J. et Infante A. (2016), « [Impact des droits de succession sur le comportement d'accumulation du patrimoine](#) », *Revue française d'économie*, vol. 31(1), juillet, p. 187-206.
18. Voir par exemple l'évaluation de Bourdon *et al.* (2014), qui réalise des enquêtes téléphoniques visant à connaître la situation des apprentis concernés par la politique publique et les raisons de leur départ de l'entreprise. Bourdon J., Guegnard C. et Michot C. (2014), « [La sécurisation des parcours des apprentis à l'épreuve du choix politique](#) », *Formation Emploi - Revue française de sciences sociales*, n° 126, avril-juin, p. 79-98.
19. Cette méthode consiste à tirer aléatoirement un groupe de traitement et un groupe de contrôle, ce qui permet de limiter les biais de sélection dans l'estimation de l'impact d'une intervention publique. Voir par exemple l'évaluation de Behaghel *et al.* (2015), qui porte sur l'expérimentation des CV anonymes, a fait l'objet de 109 citations parmi lesquelles on retrouve une citation par des décideurs, une citation par une administration publique et deux articles de presse (un article de *Capital* et une tribune dans *Libération*), le reste étant des citations de chercheurs. Behaghel L., Crépon B. et Le Barbanchon T. (2015), « [Unintended effects of anonymous résumés](#) », *American Economic Journal: Applied Economics*, vol. 7(3), juillet, p. 1-27.
20. Voir par exemple l'évaluation de Bressoux *et al.* (2009), à propos de la formation des enseignants : « The effect of class size is shown to be significant and negative: a smaller class size improves student achievement. » Bressoux P., Kramarz F. et Prost C. (2009), « [Teachers' training, class size and students' outcomes: Learning from administrative forecasting mistakes](#) », *The Economic Journal*, vol. 119(536), février, p. 540-561.
21. Voir par exemple l'évaluation de Moschion (2012), à propos de la préscolarisation : « Cet impact est significativement négatif pour les mères diplômées au maximum du baccalauréat et non significatif pour les mères plus diplômées. » Moschion J. (2012), « [Concilier vie familiale et vie professionnelle. L'effet de la préscolarisation](#) », *Revue économique*, vol. 63(2), janvier, p. 187-214.
22. Voir par exemple l'évaluation de Bunel *et al.* (2009), à propos d'un allègement de cotisations sociales en 2003 : « Au total, son effet sur l'emploi total s'avère ambigu, qu'il soit mesuré en effectif ou en équivalent temps plein. Elle a cependant contribué à la hausse des rémunérations dans les deux catégories d'entreprises. » Bunel M., Gilles F. et L'Horty Y. (2009), « [Les effets des allègements de cotisations sociales sur l'emploi et les salaires : une évaluation de la réforme de 2003](#) », *Économie et Statistique/Economics and Statistics*, n° 429-430, décembre, p. 77-105.
23. Voir par exemple l'évaluation de Vaillant *et al.* (2011), à propos du prix du tabac : « To this end, tax policy on RYO tobacco should equal the price per gram of cigarettes; currently, one gram of cigarette is worth 0.265€ while one gram of RYO tobacco is 0.13€. » Vaillant N.G., Ben lakhdar C. et Lebrun T. (2011), « [On the way of tobacco quitting: A VAR approach](#) », *Economics Bulletin*, vol. 31(4), p. 3253-3258.
24. Voir par exemple l'évaluation de Bono *et al.* (2019), à propos du dispositif Scellier : « Notre conclusion principale est donc que la politique de règlement d'urbanisme doit accompagner, voire précéder, tout dispositif incitatif à la construction de logements locatifs. » Bono P.-H. et Trannoy A. (2019), « [L'impact du dispositif Scellier sur les prix des terrains à bâtir](#) », *Économie, et Statistique/Economics and Statistics*, n° 507-508, p. 93-116.



## LES ÉVALUATIONS ACADÉMIQUES SONT FAIBLEMENT CITÉES EN DEHORS DES CHERCHEURS, MAIS LE SONT DE PLUS EN PLUS

De façon agrégée, l'analyse textuelle a relevé un total de 8 294 citations pour l'ensemble des 227 évaluations d'impact entre 2008 et 2020. Ce total correspond à une moyenne d'environ 37 citations par évaluation. Près de 94 % des citations ont été faites par des chercheurs entre eux. En d'autres termes, seulement 6 % du total des citations sont le fait des décideurs publics, des administrations publiques et des médias. Parmi ces 218 évaluations citées au moins une fois par des chercheurs, 20 % ont été citées au moins 50 fois.

Par contraste, les autres acteurs (décideurs publics, administrations publiques et médias) font beaucoup moins de citations. En particulier, les décideurs publics citent seulement 72 des 227 évaluations (soit environ 32 %) et leurs citations correspondent à uniquement 1,2 % du total des citations. Ainsi en moyenne, chaque évaluation est citée 0,45 fois par des décideurs publics pour l'ensemble des 227 évaluations. Les administrations publiques citent un peu plus les évaluations que les décideurs, en citant chaque évaluation en moyenne 1,1 fois. Les médias citent en moyenne 0,7 fois chaque évaluation. Seulement 20 % des évaluations sont citées au moins une fois par l'ensemble des acteurs. L'évaluation de Bressoux *et al.* (2009)<sup>25</sup>,

qui se sont intéressés à l'impact de la formation des professeurs et de la taille des salles de classe sur la réussite des élèves, est l'une de ces évaluations citées par tous les types d'acteurs (décideurs publics, administrations publiques, chercheurs et médias).

En croisant le nombre de citations avec les caractéristiques des évaluations, un certain nombre de faits stylisés apparaissent.

### L'année de publication de l'évaluation

Les évaluations les plus anciennes sont les plus citées (71 citations pour une évaluation de 2008 contre 12 citations pour une évaluation de 2020). Cette tendance tient avant tout au fait que les évaluations les plus récentes ont moins eu le temps d'être citées. En estimant le nombre de citations faites chaque année, pour l'ensemble des évaluations produites entre 2008 et cette année-là, il apparaît une très forte tendance à la hausse. Ainsi, près de 1 700 citations ont été faites en 2020 contre une cinquantaine en 2008. Cette tendance à la hausse tient mécaniquement au stock croissant d'évaluations disponibles (9 en 2008 et 227 en 2020). En corrigeant cet effet volume, une hausse, certes plus faible, subsiste : en 2012, étaient faites 3,9 citations par évaluation disponible, contre près de 7,4 par évaluation disponible en 2020<sup>26</sup>. Cette dynamique vaut également pour les citations faites par les décideurs, les administrations et les médias : entre 2012 et 2020, et sur les

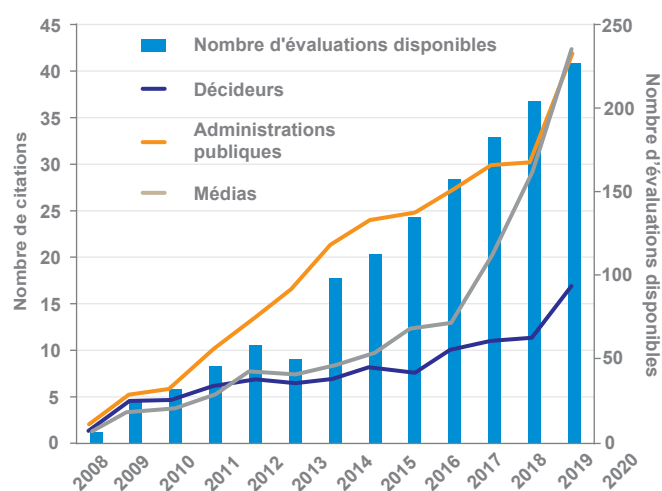
**Tableau 1 – Nombre de citations des évaluations académiques**

	Par les décideurs	Par les administrations	Par les chercheurs	Par les médias	Total	Total (hors chercheurs)
<b>Nombre de citations</b>	102	255	7 774	163	8 294	520
<b>Part dans le total des citations</b>	1,2 %	3,1 %	94 %	1,9 %	100 %	6,0 %
<b>Nombre moyen de citations par...</b>	0,45	1,1	34	0,7	37	2,3
<b>Nombre d'évaluations citées au moins 1 fois</b>	72	149	218	90	223	167

Lecture : pour l'ensemble des 227 évaluations, les décideurs ont fait un total de 102 citations, soit 1,2 % du total des 8 294 citations, et soit 0,45 citation par évaluation. Ils citent 72 des 227 évaluations. En particulier pour la dernière colonne, 167 évaluations sont citées par au moins un des types d'acteurs suivants : les décideurs, les administrations ou les médias.

Source : France Stratégie

**Nombre de citations faites annuellement 2008-2020**



Lecture : en 2020, les décideurs publics ont fait 17 citations sur les 227 évaluations disponibles.

Source : France Stratégie

25. Bressoux P., Kramarz F. et Prost C. (2009), « Teachers' training, class size and students' outcomes: learning from administrative forecasting mistakes », *The Economic Journal*, vol. 119(536), mars, p. 540-561.

26. Rappelons qu'il s'agit là d'un minorant du nombre total réel de citations, étant donné que des évaluations publiées avant 2008 peuvent également avoir été citées entre 2008 et 2020 mais ne sont pas comptabilisées ici. Ce minorant donne déjà un ordre de grandeur exact, puisque le nombre d'évaluations d'impact est relativement faible avant 2008 (moins de 15 % du nombre entre 2008 et 2020).

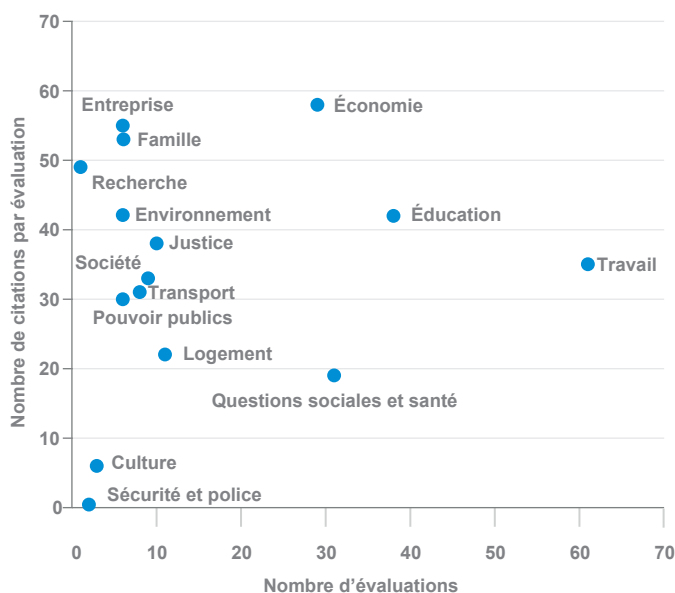
évaluations disponibles chaque année, le nombre annuel de citations faites par ces acteurs réunis passe de 28 à plus d'une centaine, soit une multiplication par 3,6. En corrigeant l'effet volume, il ressort en revanche que le nombre de citations par évaluation disponible reste relativement stable, autour de 0,4 sur la période considérée.

La production croissante d'évaluations semble donc s'accompagner d'une utilisation grandissante de ces dernières. En outre, le délai entre la publication de l'évaluation et sa citation est généralement compris entre deux et sept années, avec une moyenne à quatre années et demie.

### La langue de l'article

Les décideurs publics et les administrations publiques citent davantage les évaluations rédigées en français (respectivement 0,64 et 1,51 citation par évaluation) que celles rédigées en anglais (respectivement 0,35 et 0,91). C'est le contraire qui s'observe dans les médias, et surtout chez les chercheurs : ces derniers citent quatre fois plus les évaluations en anglais que celles en français. La langue anglaise peut être associée à des revues académiques plus prestigieuses (à l'instar de l'*American Economic Journal*), d'où l'écho plus important des articles rédigés en anglais auprès des chercheurs. En outre, les chercheurs parvenant à publier dans des revues prestigieuses anglo-saxonnes jouissent vraisemblablement d'une notoriété leur permettant de citer plus facilement leurs travaux dans les médias.

**Graphique 4 – Nombre d'évaluations et nombre de citations par évaluation, par thématique**



Lecture : chacune des 61 évaluations relevant de la thématique « Travail » a été en moyenne citée 35 fois (décideurs, administrations, chercheurs et médias réunis).

Source : France Stratégie

### La thématique de l'évaluation

La thématique « économie » figure à la fois parmi les plus évaluées et parmi celles où les évaluations sont les plus citées, avec plus de 50 citations par évaluation. Par contraste, des thématiques comme « sécurité et police » et « culture » sont relativement peu évaluées et regroupent les évaluations les moins citées (entre 0 et 6 citations par évaluation). En outre, et bien qu'elle soit de loin la plus évaluée (61 évaluations), la thématique « travail » n'est pas particulièrement citée : chacune des évaluations relevant de cette thématique est citée en moyenne 35 fois, soit proche de la moyenne générale de 37.

### La nature du producteur de l'évaluation

En excluant les citations faites par les chercheurs, il apparaît que ce sont les évaluations réalisées par les administrations publiques, seules ou en collaboration avec des chercheurs, qui sont les plus citées. Ces évaluations, qui correspondent à 32 % de l'échantillon, reçoivent chacune plus de trois citations en moyenne. Les évaluations produites par l'association de plusieurs universités ou centres de recherches (41 % de l'échantillon) sont chacune citées en moyenne 2,2 fois ; et seulement 1,3 fois pour les évaluations réalisées par des chercheurs issus de la même université ou du même centre de recherches (26 % de l'échantillon).

### La mobilisation ou non d'une analyse qualitative

Si la mobilisation d'une méthode qualitative semble *a priori* favoriser l'utilisation de l'évaluation dans le débat public (plus grande intelligibilité et richesse des résultats, implication des parties prenantes, etc.), ce n'est pas ce qui ressort de notre analyse. Du côté des décideurs et des administrations publiques réunis, le nombre de citations est en effet très proche (environ 2) en présence ou non d'une méthode qualitative. Seuls les chercheurs semblent préférer les évaluations strictement quantitatives (35 citations par évaluation contre 30 citations pour les évaluations comprenant une méthode qualitative). Cela pourrait tenir au fait que les évaluations strictement quantitatives sont publiées dans des revues plus prestigieuses.

### Le recours ou non à une méthode d'expérimentation aléatoire

En distinguant les seize évaluations reposant sur une expérimentation aléatoire des autres évaluations, il ressort que les premières sont plébiscitées auprès des décideurs et des chercheurs. En effet, les décideurs les citent près de 50 % plus et les chercheurs près de trois fois plus. Outre leurs qualités statistiques et leur intelligibilité, de telles évaluations correspondent de façon intrinsèque à des politiques dont l'expérimentation et l'évaluation suscitent un fort intérêt en soi.



### La nature conclusive des résultats de l'évaluation

S'il est compréhensible que les évaluations en demi-teinte soient moins utilisées du fait de leur manque de clarté, on s'étonne que les évaluations les plus conclusives soient les moins citées (2,1 citations hors citations entre chercheurs), et les évaluations conditionnellement conclusives les plus citées (2,7 citations hors citations entre chercheurs). Cette hiérarchie peut s'expliquer à deux titres. D'abord, les évaluations conditionnellement conclusives, par définition, comportent des résultats significatifs et d'autres non, selon les contextes et les variables retenues. De telles évaluations sont donc intrinsèquement plus riches et plus contrastées, ce qui est propice à susciter un débat public et auprès des chercheurs. Le cas de l'évaluation du CICE est emblématique : si ses résultats sont conditionnellement conclusifs, ils abordent des impacts variés (salaires, emploi, compétitivité, etc.) dans un contexte politique sensible. De plus, les évaluations conditionnellement conclusives, et à plus forte raison les évaluations non conclusives, peuvent être interprétées de façon plus libre, que ce soit pour défendre la politique ou pour la contester. Le biais du *cherry picking*, qui consiste à choisir les évaluations et les résultats qui vont dans le sens de ses inté-

rêts et de ses préjugés, pourrait donc intervenir également à la défaveur des évaluations trop conclusives.

### La présence ou non de recommandations

Si la plupart des chercheurs interrogés estiment que la formulation des recommandations ne fait pas partie du champ de l'exercice évaluatif, ces dernières semblent appréciées des décideurs pour leur côté opérationnel. De manière globale, il en ressort que les évaluations formulant des recommandations précises sont les plus délaissées, et celles formulant des recommandations générales les plus citées. Ce constat vaut en particulier pour les décideurs publics (0,37 citation par évaluation pour les premières, contre 0,57 pour les secondes), et surtout pour les chercheurs (20 contre 48 citations). Au mieux, les évaluations formulant des recommandations précises sont autant citées que les évaluations ne formulant aucune recommandation. Les décideurs publics pourraient en effet être rebutés à l'idée de se voir dicter une conduite à travers des recommandations trop précises. Les chercheurs aussi pourraient délaissier les évaluations formulant des recommandations précises en jugeant que l'exercice devient trop politique ou pas assez académique; cette interprétation est corroborée par les entretiens réalisés.

**Tableau 2 – Récapitulatif du nombre moyen de citations en fonction des caractéristiques de l'évaluation**

Caractéristiques de l'évaluation	Part des évaluations citées par l'ensemble des acteurs	Nombre moyen de citations par évaluation (hors citations chercheurs)	Nombre moyen de citations chercheurs par évaluation	Taille de l'échantillon
<b>LANGUE</b>				
Anglais	14 %	2,0	47	147
Français	23 %	2,9	11	80
<b>NATURE DU PRODUCTEUR</b>				
Administration	28 %	3,8	24	25
Même milieu académique	7 %	1,3	20	58
Mixte administration - académique	21 %	2,9	44	48
Mixte académique	19 %	0,5	41	94
Autres	0 %	0,5	20	2
<b>MÉTHODE QUALITATIVE (EN PLUS D'UNE MÉTHODE QUANTITATIVE)</b>				
Non	25 %	2,3	35	205
Oui	17 %	2,5	30	22
<b>MÉTHODE ALÉATOIRE</b>				
Non	7 %	2,3	30	211
Oui	17 %	2,6	90	16
<b>RÉSULTATS CONCLUSIFS</b>				
Oui	17 %	2,1	32	130
Conditionnellement	18 %	2,7	42	28
Non	17 %	2,6	36	69
<b>RECOMMANDATIONS</b>				
Non	14 %	2,1	29	134
Orientations générales	23 %	2,6	48	74
Recommandations explicites	16 %	2,1	20	19
<b>ÉCHANTILLON GÉNÉRAL</b>	<b>17 %</b>	<b>2,3</b>	<b>34</b>	<b>227</b>

Lecture : au regard de la caractéristique « langue de l'évaluation », il ressort que 14 % des 147 évaluations rédigées en anglais sont citées au moins une fois par l'ensemble des acteurs (décideurs, administrations, médias, chercheurs) ; en écartant les citations faites par les chercheurs entre eux, elles sont chacune citées 2,0 fois en moyenne par les autres acteurs (décideurs, administrations et médias ensemble) ; elles sont chacune citées 47 fois par les chercheurs.

Source : France Stratégie

## CONCLUSION : APPORTS ET PERSPECTIVES

En définitive, les évaluations d'impact publiées dans des revues académiques à comité de lecture sont de plus en plus citées dans le débat public, que ce soit auprès des décideurs, des administrations publiques, des médias et des chercheurs entre eux. Elles sont d'autant plus citées – et en particulier par les décideurs et les administrations publiques – qu'elles sont en français, produites par des administrations (ou en collaboration avec elles), ou qu'elles formulent des résultats riches (et pas nécessairement conclusifs) et des recommandations suffisamment larges. De façon transverse, les évaluations issues de comités mixtes d'évaluation, associant les administrations publiques à d'autres acteurs (chercheurs, société civile, etc.), sont significativement plus citées. Les évaluateurs pourraient s'inspirer de ces corrélations pour renforcer l'utilisation de leurs travaux dans le débat public. En dépit de cette dynamique, il reste que le nombre total de citations – en écartant celles faites par les chercheurs entre eux – reste globalement faible de la part des décideurs publics, des administrations publiques et des médias.

En outre, la présente étude laisse entrevoir plusieurs développements. Il peut être intéressant, à travers des études de cas, d'interroger les usages implicites de l'évaluation (ne passant pas par des citations formelles), ou leurs mésusages (une citation déformée des résultats). On pourrait aussi appréhender d'autres caractéristiques des évaluations (leur longueur, la notoriété de leurs auteurs, la formulation de leurs questions évaluatives, etc.) ou étudier des facteurs exogènes pouvant favoriser ou non l'utilisation des évaluations (le niveau d'acculturation à l'évaluation dans un service de presse ou une institution, la sensibilité politique d'un sujet, etc.). De plus, il peut être utile de tester des périodes temporelles alternatives afin d'évaluer les effets calendaires (par exemple les délais de publication d'une évaluation) qui peuvent influencer les résultats de l'analyse. De même, d'autres catégories statistiques (par exemple la liste des thématiques des champs de politique publique) peuvent être testées pour davantage appréhender les potentiels effets de structure<sup>27</sup>.

Enfin et surtout, il est à noter que l'échantillon des évaluations considérées ici renvoie à une acception restrictive de l'évaluation des politiques publiques. D'autres types d'évaluations existent : elles peuvent être publiées en dehors des revues académiques à comité de lecture ; elles peuvent être produites par des administrations, des parlementaires ou des cabinets privés ; elles peuvent porter sur d'autres critères évaluatifs que celui d'impact (mise en œuvre, coût, etc.) ; et elles peuvent reposer sur des méthodes autres que quantitatives et causales. L'utilisation de ces évaluations, entendues dans un sens large, fait précisément l'objet de la seconde note d'analyse<sup>28</sup> issue du même rapport<sup>29</sup>.

Mots clés : évaluation de politique publique, citations, débat public, décideurs publics, administrations publiques, chercheurs, médias.

27. Ces différents biais potentiels, évoqués dans le rapport, ont été en partie testés.

28. Baïz A. et al. (2022), « Quelles évaluations sont mobilisées avant et après le vote d'une loi ? », *La Note d'analyse*, n° 110, France Stratégie, juin ; Baïz A. et al. (2022), *Quelles évaluations des politiques publiques pour quelles utilisations ?*, rapport, France Stratégie, juin.

29. Pour les entretiens réalisés, nos remerciements vont aux chefs de service de l'Assemblée nationale : Thierry Anjubault (commission des Affaires économiques), Simon Corley (commission des Affaires sociales), Adrien Gaffier et Sylvain Sautier (commission des Finances), Emmanuelle Lavie (commission du Développement durable et de l'aménagement du territoire), Christophe Maisonneuve (comité d'évaluation et de contrôle des politiques publiques). Pour leurs retours, nos remerciements vont aussi à la centaine de chercheurs et d'évaluateurs interrogés, ainsi qu'à la Cour des comptes, à l'Insee, aux services statistiques ministériels sollicités et à la Direction générale du Trésor.



Directeur de la publication : Gilles de Margerie, commissaire général ;  
directeur de la rédaction : Cédric Audenis, commissaire général adjoint ;  
secrétariat de rédaction : Valérie Senné, Olivier de Broca ;  
dépôt légal : juin 2022 - N° ISSN 2556-6059 ;

contact presse : Matthias Le Fur,  
directeur du service Édition-Communication-Événements,  
01 42 75 61 37, matthias.lefur@strategie.gouv.fr

RETROUVEZ LES DERNIÈRES ACTUALITÉS DE FRANCE STRATÉGIE SUR :



Institution autonome placée auprès du Premier ministre, France Stratégie contribue à l'action publique par ses analyses et ses propositions. Elle anime le débat public et éclaire les choix collectifs sur les enjeux sociaux, économiques et environnementaux. Elle produit également des évaluations de politiques publiques à la demande du gouvernement. Les résultats de ses travaux s'adressent aux pouvoirs publics, à la société civile et aux citoyens.